

RNAhybrid: microRNA target prediction easy, fast and flexible

Jan Krüger and Marc Rehmsmeier*

Center for Biotechnology, CeBiTec, Universität Bielefeld, 33594 Bielefeld, Germany

Received February 14, 2006; Revised March 28, 2006; Accepted March 29, 2006

ABSTRACT

In the elucidation of the microRNA regulatory network, knowledge of potential targets is of highest importance. Among existing target prediction methods, RNAhybrid [M. Rehmsmeier, P. Steffen, M. Höchsmann and R. Giegerich (2004) *RNA*, 10, 1507–1517] is unique in offering a flexible online prediction. Recently, some useful features have been added, among these the possibility to disallow G:U base pairs in the seed region, and a seed-match speed-up, which accelerates the program by a factor of 8. In addition, the program can now be used as a webservice for remote calls from user-implemented programs. We demonstrate RNAhybrid's flexibility with the prediction of a non-canonical target site for *Caenorhabditis elegans* miR-241 in the 3'-untranslated region of *lin-39*. RNAhybrid is available at <http://bibiserv.techfak.uni-bielefeld.de/rnahybrid>.

INTRODUCTION

microRNAs (miRNAs) are 19–24 nt long RNAs that post-transcriptionally silence their target genes by binding to the target mRNAs (1,2). Upon near-perfect hybridization around the middle of the miRNA/target duplex, the target is cleaved and subsequently degraded. With less tight hybridizations, the target can be degraded or blocked from translation. miRNAs are key players in important cellular activities such as proliferation, morphogenesis, apoptosis and differentiation (3). Besides an investigation of the mechanistic aspects of miRNA silencing, the elucidation of the miRNA regulatory network is a major challenge. With that, knowledge of potential targets is of highest importance. For human, more than 300 miRNAs have experimental support (4), with at least 800 being suspected (5). The total number of targeted genes is estimated to be one-third of the whole human gene complement, 10 000 genes (6). In stark contrast is the current number of experimentally validated targets, which according to the Diana TarBase is 55 for human (7). For fly, the situation is slightly

better with a reported number of 75 validated targets. A number of prediction methods have contributed to a large extent in the generation of interesting hypotheses about possible miRNA/target relationships (6,8–16). Here we review RNAhybrid (16) which among these methods is unique in offering a flexible online prediction. The RNAhybrid online version is used well over 1000 times per month. In Ref. (16), it was shown that RNAhybrid predicts bona fide targets in *Drosophila melanogaster* at high specificity. Among these targets were the proapoptotic genes *grim*, *reaper* and *sickle*, where *sickle* had not been predicted previously, but was experimentally tested because of its functional context with *grim* and *reaper*. Recently, some useful features have been added to RNAhybrid, among these the possibility to disallow G:U base pairs in the seed region, and a seed-match speed-up, which accelerates the program by a factor of 8. The program can now also be used as a webservice for remote calls from user-implemented programs, thus eliminating the need for a local installation. RNAhybrid is available at <http://bibiserv.techfak.uni-bielefeld.de/rnahybrid>. Researchers who use RNAhybrid are asked to cite this article and Ref. (16).

MATERIALS AND METHODS

Algorithmic core

The algorithmic core of RNAhybrid is a variation of the classic RNA secondary structure prediction (17). Instead of a single sequence that is folded back onto itself in the energetically most favourable fashion, RNAhybrid determines the most favourable hybridization site between two sequences. Though in principle these two sequences can be arbitrarily long, for microRNA target prediction, the target candidate will be rather long (hundreds to thousands of nucleotides) and the miRNA will be between 19 and 24 nt. Since microRNA/target interactions have not been reported to contain bifurcations (also called multi-loops), these are not considered by RNAhybrid, thus considerably increasing the speed of the algorithm. RNAhybrid does not use any RNA folding or pairwise sequence alignment code, but implements an algorithm that was specifically designed for RNA hybridization [see also (16)].

*To whom correspondence should be addressed. Tel: +49 0 521 106 2905; Fax: +49 0 521 106 6411; Email: marc@techfak.uni-bielefeld.de

Features of the online version

The online version of RNAhybrid is an easy-to-use web interface in which the user can upload his or her own miRNA and candidate target sequences. A number of options give broad control over the kind of interaction the program looks for. A prevailing assumption about functional miRNA/target interactions is the necessity of a 'seed' (6), a perfect Watson-Crick match between miRNA and target at miRNA positions 2–7 or 8. However, experimentally validated miRNA/target duplexes in *Caenorhabditis elegans* appear to have unpaired nucleotides in this very seed region (18). In (11), it was

experimentally shown that a target site with a seed region as small as only 4 nt can be functional as long as there is a compensatory hybridization at the miRNA 3' end. RNAhybrid answers this heterogeneity by allowing the user to freely choose the (algorithmic) necessity and nature of a seed. First, the position and length of the seed can be defined; second, G:U wobble base pairs with the seed may be allowed or not and third, the request for a seed in the prediction can be refrained from altogether. The disallowance of G:U pairs in the seed is one of the new features and has been requested frequently. Another novelty is a 'seed-match speed-up', in which in an

Table 1. A simple client program in Perl that remotely invokes the RNAhybrid webservice

```
#!/usr/bin/env perl
use SOAP::Lite;
# -----
# define variables
# -----
my $microRNA = ">microRNA a sample microRNA sequence\n".
    "UGUGGUAGUAGGUUGUAUAGU";
my $target = ">targetRNA a sample targetRNA sequence\n".
    "GCAUGGCUGAUUUAUGCAACAAUUCUACCUCAGCCUGAGCUGCG";
my @params = ("-D","true");
my $statuscode = 601;
my $result = 601;
my $wsdl="http://bibiwsserv.techfak.uni-
bielefeld.de/RNAhybrid/axis/RNAhybridPort?wsdl";
# -----
# create a service object with fault handler
# -----
my $service = SOAP::Lite->service($wsdl)
    ->on_fault(sub{my($soap, $res) = @_;
        print $res->faultstring."\n";
        $statuscode = (split(":", $res->faultstring))[0];});
# -----
# call request_orig and abort if an error occurs
# -----
my $id = $service
    ->request_orig(\@params, $microRNA, $target);
if ($statuscode == 601) {
    print "get id : $id\n";
} else {
    exit(1);
}
# -----
# call response_orig until the result is calculated (active polling)
# -----
while($statuscode == $result && $statuscode < 700) {
    $result = $service
        ->response_orig($id);
    # wait 2 seconds before calling response_orig again
    sleep(2);
}
# -----
# print result to STDOUT
# -----
print ".$result."\n\n";
```

initial filter step, candidate targets are searched for seed matches, only upon finding such matches the complete hybridization around the seed-match is calculated. For non-G:U seeds of length 6, this implements a speed-up of a factor of 8. Another new option is to restrict possible sizes of unpaired regions, the loops. Both 'bulge loops', those with unpaired nucleotides on only one side, and 'internal loops', those with unpaired nucleotides on both sides, can be restricted in their length to user-defined values. This is especially useful in the prediction of plant miRNA targets. These targets usually exhibit only a small number of unpaired nucleotides, if any (8). Restricting loop sizes to, for example 1 nt, avoids the generation of spurious hits that do not conform to established miRNA/target hybridization rules in plants. Two other useful options are the number of target sites per miRNA and target candidate the program looks for, and a threshold for the minimum free energy of the hybridization, only below which target sites are reported. This latter option is the only option that is offered by Diana microT (15), in turn the only method besides RNAhybrid that is available for online miRNA target prediction in animals. The program miRU (19) is available as an online tool, but is geared towards prediction of potential targets in plants.

RNAhybrid webservice

A new technology for invoking programs on remote computers are webservices. Providing access using webservices makes it possible to use the programs from local computers and compute results remotely without technical knowledge of the programs. In addition to the traditional browser/HTML-based web-interface, we also offer a webservice interface for using RNAhybrid in a batch job environment or from another, user-developed program. All options of the traditional submission form are supported by the webservice version. The RNAhybrid webservice is asynchronous and implements the request and response with a polling technique (Adams, H., Asynchronous operations and webservice, <http://www-106.ibm.com/developerworks/library/wsasync1/>) that follows the HOBIT standard for exchanging status information between client and server (HOBIT, Helmholtz Open Bioinformatics Technology, <http://hobit.sourceforge.net>). Users can create their own webservices client by using a webservice framework. Well-known webservice frameworks are SOAP::Lite (Perl) (a simple and lightweight interface to SOAP, <http://soaplite.com/>), AXIS (webservice framework for Java and C/C++, <http://ws.apache.org/axis>), gSOAP (C/C++ webservices and clients, <http://gsoap2.sourceforge.net>) and .NET (Microsoft .Net framework, <http://www.microsoft.com/net/>). Sample clients for Perl/SOAP::Lite and Java/Axis are available on the RNAhybrid homepage. A simple client that uses the Perl programming language with SOAP::Lite is shown in Table 1.

RESULTS

In *C.elegans*, members of the *let-7* family of miRNAs, which comprises the miRNAs *let-7*, *miR-48*, *miR-84* and *miR-241*, function in combination to affect early and late developmental timing decisions (20). *miR-48*, *miR-84* and *miR-241* control the L2-to-L3 transition, probably by binding to the *hbl-1* 3'-untranslated region (3'-UTR). *lin-41*, which acts redundantly

Table 2. ClustalW alignment of *C. elegans let-7* family sequences

cel-miR-48	UGAGGUAGGCUCAGUAGAUGCGA
cel-miR-241	UGAGGUAGGUGCGAGAAUGA--
cel-let-7	UGAGGUAG-UAGGUUGUAUAGUU
cel-miR-84	UGAGGUAG-UAUGUAAUUAUGUA
	***** **

Asterisks denote conserved positions.

with *hbl-1* in the regulation of the L4-to-adult transition, is repressed by *let-7*, but probably not by *miR-48*, *miR-84* and *miR-241*. It is suggested in (20) and has been so before in (11) that the target specificity of these miRNAs might be defined by their 3'-sequence. While the 5'-sequence from nucleotides 1 to 8 is identical in all four miRNAs, the 3'-part exhibits strong sequence diversity (see Table 2). Since *lin-41* lacks binding sites for the 5'-seed, the existence of *let-7/lin-41* target sites does not automatically give rise to *miR-48*, *miR-84* and *miR-241* sites. In fact, *let-7* is the only probable regulator of *lin-41*, and this regulation is mediated by an extended 3'-complementarity (18). In Ref. (20), Ambros and colleagues speculate that there might be genes that are specifically targeted by *miR-48*, *miR-84* or *miR-241*. To test this hypothesis, we analysed the 3'-UTRs of 33 *lin* (abnormal cell LINEage) genes, downloaded from the Ensembl database (<http://www.ensembl.org>), for target sites with extended 3'-complementarity to the members of the *let-7* family. 3'-complementarity was enforced by requiring a 'seed' from nucleotides 12 to 18, not allowing G:U base pairs. In addition to the expected *let-7/lin-41* target sites, we found a strong hit for *miR-241* in the *lin-39* 3'-UTR (see Table 3). *lin-39* encodes a homeodomain protein homologous to the Deformed and Sex combs reduced family of homeodomain proteins and is required for the specification of, among others, vulval precursor cells (21). A weaker match (data not shown) for *miR-241* was found in the 3'-UTR of *lin-45*, which is required for, among others, the induction of vulval cell fates (22). In *Caenorhabditis briggsae*, we predict a potential binding site for *miR-241* in *lin-39*, though not in the same position and of a weaker quality than in *C.elegans* (see also Table 3).

DISCUSSION

RNAhybrid is a tool for the easy, fast and flexible prediction of microRNA targets. Besides Diana microT and miRU, it is the only method available as an online tool. At the same time, RNAhybrid offers a larger choice of options and applications. As an example, we analysed the 3'-UTRs of 33 *C.elegans lin* (abnormal cell LINEage) genes for potential non-canonical target sites for any of the four *C.elegans let-7* microRNA family members. Extensive 3'-pairing was enforced by requiring RNAhybrid to form 'seed' matches at nucleotides 12–18 in the miRNA, disallowing G:U basepairs. The analysis resulted in the prediction of *lin-39* as a strong target candidate for *miR-241*. This finding supports suggestions about target specificity that is defined by 3'-complementarity (11,20). The unusual choice of the 'seed' position demonstrates RNAhybrid's flexibility. While the classic seed assumption (nucleotides 2–7 or 8) increases the statistical significance of target predictions in genome-wide analyses [see eg. (6,11,16)], one might miss bona fide target sites that do not

Table 3. Predicted target sites for *miR-241* in the *lin-39* 3'-UTR of *C.elegans* (left) and *C.briggsae* (right). The *p*-values were calculated with the download-version of RNAhybrid

target: cel-lin-39_3UTR	target: cbr-lin-39_3UTR
length: 448	length: 218
miRNA : cel-miR-241	miRNA : cbr-miR-241
length: 21	length: 21
mfe: -30.2 kcal/mol	mfe: -19.7 kcal/mol
p-value: 1.164136e-03	p-value: 6.316332e-01
position 67	position 1
target 5' A C U U 3'	target 5' UAUA U 3'
CAUUUCUCGC CCU CCUU	AUUUCUU GCUUGCUIIU
GUAAGAGCG GGA GGAG	UAAAGAG UGGAUGGAG
miRNA 3' A U U U 5'	miRNA 3' AG UG U 5'

For *C.elegans*, target sites were required to have a 'seed' from nucleotides 12 to 18 in the miRNA, and no G:U basepairs in the seed were allowed.

show this seed, as is already suggested by *let-7* and *lin-4* target sites in *C.elegans* which have bulging nucleotides in the seed region (18). Also, Stark *et al.* (11) have experimentally demonstrated that short 'seeds' of 4 nt can be compensated by 3'-complementarity. In fact, *lin-39* has not been predicted as a target of *miR-241* by any of the standard target prediction approaches [PicTar (13), TargetScanS (6), MiRanda (4)], presumably because these methods, to various extents, rely on the presence of classic seed matches (miRanda does this indirectly by favouring 5'-matching). It should be fruitful in the future to perform genome-wide predictions of non-canonical target sites.

ACKNOWLEDGEMENTS

The authors thank Carsten Drepper and Robert Heinen for valuable comments of the RNAhybrid software. J.K. and M.R. were supported by the Deutsche Forschungsgemeinschaft, Bioinformatics Initiative. Funding to pay the Open Access publication charges for this article was provided by Deutsche Forschungsgemeinschaft.

Conflict of interest statement. None declared.

REFERENCES

- Ambros, V. (2001) microRNAs: Tiny Regulators with Great Potential. *Cell*, **107**, 823–826.
- Bartel, D.P. (2004) MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*, **116**, 281–297.
- Carthew, R.W. (2006) Gene regulation by microRNAs. *Curr. Opin. Genet. Dev.*, **16**, 203–208.
- Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. and Enright, A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.
- Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E. *et al.* (2005) Identification of hundreds of conserved and nonconserved human microRNAs. *Nature Genet.*, **37**, 766–770.
- Lewis, B.P., Burge, C.B. and Bartel, D.P. (2005) Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell*, **120**, 15–20.
- Sethupathy, P., Corda, B. and Hatzigeorgiou, A.G. (2006) TarBase: a comprehensive database of experimentally supported animal microRNA targets. *RNA*, **12**, 192–197.
- Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B. and Bartel, D.P. (2002) Prediction of plant microRNA targets. *Cell*, **110**, 513–520.
- Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. and Burge, C.B. (2003) Prediction of mammalian microRNA targets. *Cell*, **115**, 787–798.
- Stark, A., Brennecke, J., Russel, R.B. and Cohen, S.M. (2003) Identification of *Drosophila* MicroRNA Targets. *PLoS Biol.*, **1**, e60.
- Brennecke, J., Stark, A., Russell, R.B. and Cohen, S.M. (2005) Principles of MicroRNA–Target Recognition. *PLoS Biol.*, **3**, e85.
- Rajewsky, N. and Sock, N.D. (2004) Computational identification of microRNA targets. *Dev. Biol.*, **267**, 529–535.
- Lall, S., Grün, D., Krek, A., Chen, K., Wang, Y.L., Dewey, C.N., Sood, P., Colombo, T., Bray, N., Macmenamin, P. *et al.* (2006) A Genome-Wide Map of Conserved MicroRNA Targets in *C. elegans*. *Curr. Biol.*, **16**, 460–471.
- John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C. and Marks, D.S. (2004) Human MicroRNA targets. *PLoS Biol.*, **2**, e363.
- Kiriakidou, M., Nelson, P.T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z. and Hatzigeorgiou, A. (2004) A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.*, **18**, 1165–1178.
- Rehmsmeier, M., Steffen, P., Höchsmann, M. and Giegerich, R. (2004) Fast and effective prediction of microRNA/target duplexes. *RNA*, **10**, 1507–1517.
- Zuker, M. and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.*, **9**, 133–148.
- Grosshans, H. and Slack, F.J. (2002) Micro-RNAs: small is plentiful. *J. Cell Biol.*, **156**, 17–21.
- Zhang, Y. (2005) miRU: an automated plant miRNA target prediction server. *Nucleic Acids Res.*, **33** Web Server issue, W701–W704.
- Abbott, A.L., Alvarez-Saavedra, E., Miska, E., Lau, N.C., Bartel, D.P., Horvitz, H.R. and Ambros, V. (2005) The *let-7* MicroRNA Family Members *mir-48*, *mir-84*, and *mir-241* Function Together to Regulate Developmental Timing in *Caenorhabditis elegans*. *Dev. Cell*, **9**, 403–414.
- Burglin, T.R. and Ruvkun, G. (1993) The *Caenorhabditis elegans* homeobox gene cluster. *Curr. Opin. Genet. Dev.*, **3**, 615–620.
- Hsu, V., Zobel, C.L., Lambie, E.J., Schedl, T. and Kornfeld, K. (2002) *Caenorhabditis elegans lin-45 raf* is Essential for Larval Viability, Fertility and the Induction of Vulval Cell Fates. *Genetics*, **160**, 481–492.